

Ruben Bousbib

+33 6 72 06 89 86 ✉ ruben.bousbib@gmail.com
in ruben-bousbib 📄 rubenbsb.github.io 🌐 RubenBSB

EDUCATION

M.Res., École Normale Supérieure Paris-Saclay **Paris, France**
MVA program: Mathematics, Vision, Learning 2020 - 2021
Classes: Graphs in ML, Convex Optimization, Reinforcement Learning, Advanced Deep Learning, Medical Imaging, Computational Statistics. *Summa Cum Laude*

M.Sc., Paris-Saclay University - CentraleSupélec **Paris, France**
Applied Mathematics and Computer Science 2017 - 2021
Classes: Advanced Statistics and Probabilities, Natural Language Processing, Random Signal Processing, Advanced Machine Learning. Ranked 8/500 - GPA: 3.96/4

Lycée Pasteur **Paris, France**
Preparatory classes 2014 - 2017
Intensive preparation courses in Maths, Physics and Computer Science for the highly competitive entrance exam to the French "Grandes Écoles".

EXPERIENCE

Deep Learning Researcher **Remote**
AssemblyAI Apr. 2022 - Now
Large-scale speech processing and audio intelligence.

- Implemented a scalable multi-node distributed training pipeline for GPUs and TPUs on multiple platforms (GKE, VertexAI and SLURM cluster).
- Built state-of-the-art multi-lang ASR models and created ready-to-deploy packages out of it.
- Implemented toolkits for profiling, benchmarking and optimizing models automatically using TensorRT-LLM and ONNX.

Research Engineer Intern **Paris, France**
Meta AI Aug. - Dec. 2021
NLP-inspired software fuzzing - Supervised by Lucas Hosseini and Edouard Grave.

- Developed a fuzzing dataset generation tool to collect interesting mutations of compiled programs.
- Implemented from scratch a distributed training pipeline and Transformer-based models to learn byte-level code coverage information.

Research Intern **Paris, France**
Sonos, Inc. May - Aug. 2021
Spoken Language Understanding – Supervised by Prof. Yannick Estève and Alice Coucke.

- Developed multimodal models and ML experiments for Addressee Detection on privacy-aware and embedded voice assistance technologies.
- Implemented the data generation pipeline and deployed to production.

Research Intern **Paris, France**
École Normale Supérieure - Cognitive Machine Learning team Sep. 2019 - Mar. 2020
Speech processing - Supervised by Prof. Emmanuel Dupoux (Meta AI).

- Implemented an approach for domain adaptation of end-to-end voice activity detection.
- Developed deep multi-task models for voice type classification from the raw waveform.
- Designed and run the experiments on a computer cluster and published results in top conference.

Software Engineer Intern **Paris, France**
Sopra Steria Jun. - Aug. 2018

- Developed a software in Java to analyse the payroll system of the French armed forces.
- Implemented bash scripts to monitor data warehousing.

PUBLICATIONS

📄 M. Lavechin, **R. Bousbib**, H. Bredin, E. Dupoux, A. Cristia. "An open-source voice type classifier for child-centered daylong recordings", in *Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2020.

📄 M. Lavechin, M-P. Gill, **R. Bousbib**, H. Bredin, L. Paola García-Perera. "End-to-end Domain-Adversarial Voice Activity Detection", in *Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2020.

Submitted to INTERSPEECH 2024:

📄 K. Zhang, L. Chkhetiani, F. M. Ramirez, Y. Khare, A. Vanzo, M. Liang, S. R. Martin, G. Oexle, **R. Bousbib**, T. Peyash, M. Nguyen, D. Pulliam, D. Donato. "Conformer-1: Robust ASR via Large-Scale Semisupervised Bootstrapping".

SKILLS

Programming: Python, C/C++, CUDA, Java, Matlab, shell

ML: PyTorch, JAX, Tensorflow, Triton

Engineering: git, docker, distributed systems, SLURM, Spark, AWS, GCP, Ray.io

Miscellaneous: JavaScript, NodeJS, ReactJS, LaTeX

PROJECTS

BreastMIL:

Attention-based Multiple Instance Learning for breast cancer metastases prediction from very high resolution histology images.

DisasterNLP:

Developed a disaster tweets classification system with pre-trained BERT architecture.

Link: 📄 Report

TumorDetection:

Performed tumors detection from mammography MRI images with MIAS Dataset and using a Faster R-CNN architecture.

Link: 🖼️ Poster

DeepOCR:

Built from scratch a handwritten text recognition system in PyTorch using a CNN-LSTM-CTC architecture, under the supervision of Prof. Renaud Séguier (CentraleSupélec).

Links: 🌐 Github 🖼️ Poster

SubwayLinker:

Implemented a link prediction algorithm to optimize the building of a subway network under a cost constraint.

Link: 📄 Report

COMMUNITY SERVICE

AI Junior Manager:

Oct. 2017 - May 2019

Junior CentraleSupélec is an association where student can undertake technical missions for companies.

- Management of artificial intelligence projects.
- Recruitment of students.
- Price estimation, quality assurance and legislative control.

LANGUAGES

French: Native

English: Fluent

Hebrew: Conversational

INTERESTS

DJing: soundcloud.com/isaac-bousbib 📄

Cinema

Piano